The Causal Effect of ACA Subsidies on Insurance Coverage Status Among California Adults

---

William Vereyken

December 7th, 2022

Abstract

California began offering its citizens access to subsidies for health insurance per the Affordable Care Act (ACA) in late 2013, with coverage beginning January 1st, 2014. The goal of these subsidies was to increase insurance coverage while decreasing the cost burden of health insurance for low and medium-low income families. This paper seeks to assess the causal effects of these subsidies on individuals' health insurance coverage statuses using two difference in differences models and a parametric and nonparametric regression discontinuity model. Leveraging the California Health Information Surveys from 2011 through 2015, this paper provides results for California's overall adult population as well as California's African American adult population so they may be assessed individually and compared.

Keywords: Health Insurance Subsidy, Affordable Care Act, Insurance Coverage, Difference in Differences, Sharp Regression Discontinuity, Nonparametric Regression Discontinuity, Local Linear Regression

I - Introduction

This paper investigates the causal effect of ACA healthcare subsidies on California adults' healthcare insurance coverage status. The original hypothesis that drove this paper was the same one that drove the creation of ACA subsidies - subsidizing an individual's healthcare causes them to be more likely to purchase healthcare insurance. This stems from the idea that the decreased cost of health insurance and increased effective household income due to the subsidy makes people purchase more insurance (while it is not a main point of this paper, this underlying assumption would indicate that health insurance is a normal and/or ordinary good). In addition to this, this paper aims to empower comparison of the effects of subsidies on the overall adult population versus the effects on African American adult population by running empirical models on one set of data that includes all races and one set of data that is restricted to African American individuals. This allows for the results of these models to be compared to see if the estimated effects differ for the overall population versus the African American population.

Identifying the causal effects of healthcare subsidies on insurance coverage status, as well as discrepancies in these effects across different races, is important for many reasons. For example, given the social and economic benefits of health care and health insurance, motivating increased coverage is a frequent goal for legislators. Subsidies are one such way to do this, however, they can be extremely costly to the government and taxpayers. Therefore, understanding how subsidies directly affect healthcare insurance is critical in forming effective, efficient legislation. Similar arguments can be made for understanding the difference in subsidy effects across races as legislators need to ensure that effective legislation serves all populations, not just those that are currently represented and served relatively well by the government. Another example of why this paper is important stems from a public health argument: people without insurance are known to reduce healthcare use, which can be harmful to their health and the public's health if, for example, they do not receive screenings for communicable diseases. Other negative results of uninsurance are widely promoted and discussed in health and economic literature; one example of a popular paper on this topic that discussed the numerous negative repercussion of uninsurance is "Health Consequences of Uninsurance Among Adults in the United States: Recent Evidence and Implications" by J. Michael McWilliams of Harvard's Department of Healthcare Policy (McWilliams 2009). Understanding how subsidies affect health insurance coverage statuses can allow legislators to ensure that they are properly using subsidies to combat issues from uninsurance. These are only a few examples of the importance of this topic; the ever-present presence of political and financial debate regarding health insurance and subsidies alone serves as sufficient motivation for an investigation into this topic.

There exists a plethora of relevant literature for the topic that this paper explores; however, nothing is exactly the same. For example, among other relevant literature, there are papers that research similar topics but use different models and/or datasets, as well as papers that research similar topics such as the effects of other ACA measures on insurance rates. That said, this paper aims to fill a particular gap in the current literature - assessing the causal effect of healthcare subsidies on individuals' expected health insurance coverage statuses, as well as potential racial disparities of this effect, by leveraging difference in differences and regression discontinuities models on a robust panel dataset made from merged California Health Information Survey datasets (CHIS).

As is described in greater detail in subsequent sections, this paper aims to prove the causal effect of ACA subsidies on adult individuals' health insurance coverage statuses. To do this, multiple models are employed: a difference in differences model that is run on two different sets of time periods, as well as a parametric and a nonparametric regression discontinuity model.

II - Review of Relevant Literature

The idea of researching the effects of subsidies for health insurance on individuals' coverage status and decisions is certainly not novel. For this paper's topic and methodology, the relevant literature is quite expansive and includes, among many others, two main categories: papers that research health insurance subsidies effects' with different methods on different datasets and papers that use the same methods as this one but on different datasets. Similar to this paper, literature that compares the effects of health insurance subsidies on different racial groups exists across these two categories.

The first group of relevant literature is formed by papers that leverage different methods and different datasets compared to this paper but are researching similar topics. One paper that fits into this category explores subsidies and how they affect the demand for health insurance (Marquis et al. 2004). In this paper, Marquis et al. use data from the Current Population Survey, the Survey of Income and Program Participation, and the National Health Information Survey, along with a logit model to estimate the drivers of group insurance purchasing for households that do not have access to group insurance. Using this data and models, Marquis et al. find that the elasticity of demand for insurance for individuals without access to group insurance is approximately -.2 to -.4, implying that subsidies increase insurance enrollment significantly. However, Marquis et al.'s paper is different from this paper in that Marquis et al.'s results are limited to those without access to group insurance and it employs different datasets and methods.

Another paper that fits into this category of different methods and datasets investigates the three main components of the ACA directed at reducing uninsurance rates, including private insurance subsidies (Frean et al. 2017). The data used in Frean et al,'s paper is from the 2012-2015 American Community Survey and is an extremely large, panel dataset formed from merged cross-sectional data. The model employed is similar to the one employed in this paper but is a triple differences model, instead of a simple difference in differences, and is focused on analyzing the percentage of the total enrollment effects created by each of the three components of the ACA. Ultimately, Frean et al. found that roughly 40% of the reduction in the uninsured rate from 2014 to 2015 was due to private insurance subsidies. Frean et al.'s paper differs from this one in that it focuses on the splits of the total effects of the ACA policies, uses a different dataset, and uses an empirical setup that accounts for geography.

Together, these papers represent the general consensus of the broader literature with different methods and datasets from those employed in this paper: the price elasticity of demand for health insurance appears to be significant and subsidies reduce uninsurance in a causal manner. The magnitude of the causal effect, however, is debated but appears to be significant as well.

The second category of relevant literature includes papers that use the same methods as this paper, but on different datasets. One such paper investigates how a private health insurance subsidy affected the demand for private health insurance in Australia (Kettlewell et al. 2017). By leveraging a regression discontinuity model and data that includes the discontinuous increases in subsidies at ages 65 and 70, Kettlewell et al. determine that the increased subsidies have minimal effect on the decision to purchase health insurance. While this may seem to serve as evidence against the causal effect of healthcare subsidies on individuals' health insurance coverage statuses, it is important to note that the results are restricted to the Australian elderly, many of whom likely already receive a subsidy. Rather than serving as strong evidence that health insurance subsidies do not affect individuals' expected health insurance coverage statuses, these results are mainly significant in showing that there are likely differences in the effects of subsidies depending on the sizes of the subsidies and the group to which they are offered (similar to the secondary goal in this paper to investigate if there racial disparities in ACA subsidies' effects). Kettlewell et al.'s paper differs most saliently from this one in that it uses a different dataset which heavily skews for whom the results have meaning. Also, while Kettlewell et al. use a regression discontinuity which is a method also used in this paper, they do not leverage a difference in differences as this paper does.

Another paper that falls within the category of sharing an experimental method but not the leveraged dataset with this paper is "Effects Of The ACA's Health Insurance Marketplaces On The Previously Uninsured: A Quasi-Experimental Analysis" by Anna L. Goldman, et al. This paper attempts to identify the effects of the ACA's health insurance markets (which includes the effect of the subsidies) on health insurance status, access to care, and several other health economics outcomes by leveraging a difference in differences run on a similar data setup as the one employed in this paper. Goldman et al. use different data than this paper, however, as they employ data from the Medical Expenditure Panel Survey which is, as the name suggests, panel data. Ultimately, Goldman et al. find that the implementation of ACA health insurance markets significantly decreased the uninsurance rate. Goldman et al.'s paper differs from this one as it does not leverage a regression discontinuity, uses a different dataset, and does not look for racial disparities.

Another paper in this category is "Changes in Racial and Ethnic Disparities in Access to Care and Health Among US Adults at Age 65 Years" by Jacob Wallace, et al. This paper, while not focused on the direct effect of subsidies, is similar to this paper in that it evaluates how eligibility for a reduced-price health insurance offering (Medicare) affects racial disparities in health insurance coverage status (Wallace et al 2021). Wallace et al. use data from the Behavioral Risk Factor Surveillance System and the US Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research Data from 2008 to 2018, which form a cross-sectional dataset and empower regression discontinuity analysis. Ultimately, Wallace et al. concluded that access to reduced-priced health insurance increases insurance coverage rates for all racial categories analyzed, but the most significant increase was among African Americans (9.5% increase) relative to a more modest increase for White respondents (6.5%). Wallace et al.'s paper differs from this one in that it is focused on access to Medicare, not subsidies, uses a different dataset, and does not employ a difference in differences method.

Overall, the examined literature in this second category (sharing at least one experimental method with this paper but using a different dataset) agree with the literature in the first category in that they show a positive effect of health insurance subsidies on individuals' and groups' expected health insurance coverage statuses. They also show that there are disparities in the effect of subsidies depending upon to whom they are offered.

All in all, the current literature is expansive and predominantly points towards health insurance subsidies casually increasing the rates of health insurance coverage and thus the expected health insurance coverage status of individuals. Additionally, the literature points toward disparities in the strength of these effects depending on the group to which the subsidy is offered. These beliefs are supported by numerous other works not mentioned in this literature review, such as works by Mark Duggan, Charles Courtemanche, and Thomas C. Buchmueller, among numerous others (Duggan et al 2021; Courtemanche et al. 2016; Buchmueller et al. 2016).

However, none of the mentioned works combine all of this paper's attributes - regression discontinuity and difference in differences on the CHIS dataset to evaluate the causal effect of healthcare subsidies on individuals' coverage statuses with an additional comparison of the effects on African Americans and the overall population. That is where this paper fits into the current body of economic literature - as a unique combination of approach, data, and goal to hopefully increase the collective economic understanding of how subsidies affect health insurance coverage statuses, as well as bolster the existing empirical support for these shared understandings.

III - Empirical Models and Data for the Overall Adult Population

The empirical analysis in this paper aims to prove the causal effect of ACA healthcare subsidies on California adults' healthcare insurance coverage statuses. It does this for both the overall adult population and the African American adult population. The initial hypothesis is that offering subsidies to an individual causally increases their expected insurance coverage status.

This paper uses panel data constructed from several cross-sectional datasets. Specifically, the 2011, 2012, 2013, 2014, and 2015 California Health Information Survey (CHIS) datasets are merged to create a comprehensive, panel-data database that empowers analysis across individuals over time. The CHIS data is collected via a web survey and a telephone survey and has a response rate typically in the high single digits or low teens, so volunteer bias must remain a consideration while analyzing this paper's results.

For this paper, all of the individuals are adults from California. Any subgroups analyzed in this paper are a subset of the overall group of Californian adults (e.g. the African Americans subgroup is composed of adult Californians who are African American).

As the goal of the analysis is to evaluate the effect of healthcare subsidies, the data includes people with incomes of 138% and higher of the Federal Poverty Level (FPL). This eliminates the group of people with income of less than 138% of the FPL since there were other significant initiatives enacted to increase their health insurance enrollment at the same time as the subsidies which would create significant confounding in the results. In particular, the ACA's mandatory expansion of Medicaid to all people with income of less than 138% of the FPL would cause significant confounding if this data were not excluded from the dataset and ergo analysis.

The samples sizes for the cleaned (post-removal of the <138% FPL group) dataset is sufficiently large, as is shown by the following:

Figure 1: Sample Sizes Per Year for the Overall Adult Population

	Years	Overall Counts	Overall Treatment Counts	Overall Control Counts
0	2011	17355	7712	9643
1	2012	14605	7031	7574
2	2013	16779	7397	9382
3	2014	15353	6926	8427
4	2015	16368	7022	9346

For all of the following models, the outcome variable of interest is an individual's health insurance coverage status, which is a boolean that takes value 1 if the individual has health insurance and 0 if they do not. The independent variables analyzed are described below and vary for each model but include individuals' eligibility for health insurance subsidies (based on their incomes), the year, individuals' incomes, the difference between individuals' incomes and the 400% FPL cutoff for subsidy eligibility, and interaction terms.

This paper's main empirical methods of analysis are difference in differences and sharp regression discontinuity. Two difference in differences models are employed with (1) being run on the dataset with all years (2011-2015) and (2) being run on only the year of the treatment and the year prior (2013-2014). For both difference in differences models, the group of people eligible for the subsidy (which is represented by people with income between 138% and 400% FPL) form the treatment group and people not eligible for the subsidy (income over 400% FPL) are the control. These difference in differences methods are based on the work of David Card and Alan Krueger in their paper "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania" (Card and Krueger 1994). Two sharp regression discontinuities are run as well with (3) being a standard, linear regression discontinuity model and (4) being a nonparametric model that employs local linear regression. Below is a description of these models in more granular detail:

Model (1): $\mathbf{Y_{it} = B1 + B2*Subsidy_i + B3*Year2014_t + B4*(Subsidy*Year2014)_{it} + e_{it}}$

Y_{it} is whether or not they are insured (binary variable equal to 1 if the individual is insured and equal to 0 if they are not). $Subsidy_i$ is a dummy variable that takes value 1 if the individual is in the range of incomes from 138 to 400% of the FPL and 0 otherwise. It is worth noting that $Subsidy_i$ is not whether or not the individual leverages a subsidy or if a subsidy is available to them in actuality; rather, $Subsidy_i$ represents if their income would make them eligible for a subsidy. $Subsidy_i$ can be thought of as a dummy variable indicating an individual's status in the treatment (1) or control (0) group. $Year2014_t$ is a dummy variable that takes value 1 if the response is occurring in or after 2014 (since the subsidies started being offered at that time). $Year2014_t$ can be thought of as a dummy variable indicating whether or not the individual was surveyed in a year where the treatment (subsidy) was available for eligible individuals.

$(Subsidy*Year2014)_{it}$ represents the interaction term, which indicates that the individual is eligible for the subsidy in a year when the subsidy was offered. Thus, $(Subsidy*Year2014)_{it}$ is our main variable of interest as its coefficient, B4, estimates the treatment effect - the causal effect of the ACA subsidies on an individual's insurance coverage status. e_{it} is the error term. For this analysis, all data in the previously detailed dataset was employed.

Figure 2: Summary Statistics for DID Model (1) on Overall Adult Data

	count	mean	std	min	25%	50%	75%	max
Y_{it}	80460.0	0.927405	0.259472	0.000000	1.000000	1.000000	1.000000	1.000000
$Subsidy_i$	80460.0	0.448521	0.497346	0.000000	0.000000	0.000000	1.000000	1.000000
$Year2014_t$	80460.0	0.394246	0.488691	0.000000	0.000000	0.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	80460.0	0.173353	0.378555	0.000000	0.000000	0.000000	0.000000	1.000000
$Income_i$	80460.0	5.606879	4.152812	1.381042	2.610966	4.384787	7.018658	27.548209
Δ_i	80460.0	1.606879	4.152812	-2.618958	-1.389034	0.384787	3.018658	23.548209
$Subsidy_i * \Delta_i$	80460.0	-0.649051	0.874132	-2.618958	-1.389034	0.000000	0.000000	0.000000

Model (2): Same equation as (1) but run on a dataset restricted to the treatment year (2014) and the year prior (2013).

This model employs the same equation, variable definitions, causal effect estimator (B4), and key assumptions as model (1).

Figure 3: Summary Statistics for DID Model (2) on Overall Adult Data

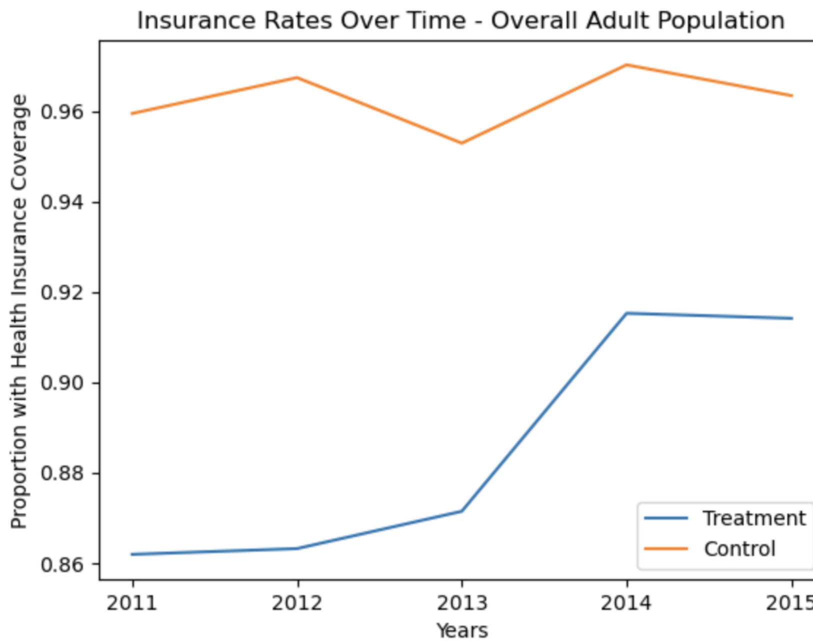
	count	mean	std	min	25%	50%	75%	max
Y_{it}	32132.0	0.930568	0.254192	0.000000	1.000000	1.000000	1.000000	1.000000
$Subsidy_i$	32132.0	0.445755	0.497057	0.000000	0.000000	0.000000	1.000000	1.000000
$Year2014_t$	32132.0	0.477810	0.499515	0.000000	0.000000	0.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	32132.0	0.215548	0.411209	0.000000	0.000000	0.000000	0.000000	1.000000
$Income_i$	32132.0	5.649200	4.175856	1.382488	2.610966	4.438642	7.092199	26.109661
Δ_i	32132.0	1.649200	4.175856	-2.617512	-1.389034	0.438642	3.092199	22.109661
$Subsidy_i * \Delta_i$	32132.0	-0.641035	0.871383	-2.617512	-1.389034	0.000000	0.000000	0.000000

Both (1) and (2), as difference in differences models, have key assumptions that must be met in order for their B4s to estimate the true treatment effect. The most notable ones for this paper are as follows:

Assumption 1: *"The Parallel Trends Assumption". If the treatment had not been introduced, the difference between the treatment and control groups would not change over time.*

This assumption essentially states that no other treatments/exogenous shocks begin affecting the treatment group at the same time as the treatment. Without this assumption being valid, significant confounding is possible. For this paper, it is highly likely that other shocks affected the treatment group simultaneously with the implementation of subsidies since the ACA legislation was implemented at the same time as many other pieces of legislation; however, their effects should be minimal based on the design of the study and dataset leveraged. To confirm this, we assess the pretends of the treatment and control group to see if they had parallel trends before the treatment is applied since we cannot observe the treatment group if it did not receive the treatment (the treatment group without the effects of the exogenous shock/treatment after the time at which the shock occurs is a counterfactual). The pretrends are shown below in Figure 4, appear to be similar, and tepidly support this assumption being valid for this paper's models on the overall adult population.

Figure 4: Percentage of Insurance Coverage Per Year - Overall Population



Note: for interpretability, the graph shows aggregate insurance rates, not an individual's insurance status. While this is different from individual insurance coverage status, the boolean nature of the status lends well to aggregation, making the visualization a quality representation of the pretrends.

Assumption 2: *"SUTVA - Stable Unit Treatment Value Assumption." Individuals respond only to their treatment status (not the treatment status of others) and there are no hidden variations in the treatment.*

This assumption states that each individual's treatment is constant and independent of the other treatments (subsidies). For this paper, while not fully correct, it is assumed that one person's subsidy receipt does not affect other individuals and their receipt of a subsidy. It could be claimed that people may communicate with others if they receive the subsidy and enjoy its benefit, leading to other people manipulating their income to receive the subsidy. Given the private nature of personal finances in California, this paper believes this sort of spillover is negligible. The second component of SUTVA dictates that the treatment does not have hidden variations. Since this paper investigates the treatment of having access to a healthcare subsidy, there is no variation - the treatment takes one of two forms: has access (1) or does not have access (0). However, if one were to be estimating how the size of the subsidy affects health insurance coverage, this would need to be taken into account. Therefore, for the purposes of this paper, it can be assumed that SUTVA is met or close enough to being met that the repercussions are negligible.

Model (3): $\mathbf{Y_i = B1 + B2{*}Subsidy_i + B3{*}Income_i + e_i}$

Y_i is whether or not the individual is insured (same as Y_{it}, but does not depend on time since that is not part of this model). $Subsidy_i$ is a dummy variable that takes value 1 if the individual is eligible for ACA subsidies and 0 if they are not (same as in (1) and (2)). $Income_i$ is the individual's household annual income, expressed as a proportion of the federal poverty level. e_{it} is the error term. Assuming the assumptions of a sharp, linear regression discontinuity hold, the causal treatment effect is estimated by B2. For this analysis, only data from the year 2015 was used so the effects of the subsidy had permeated the population fully.

Figure 5: Summary Statistics for Parametric Regression Discontinuity Model (3) on Overall Population Data

	count	mean	std	min	25%	50%	75%	max
Y_i	16368.0	0.942265	0.233248	0.000000	1.000000	1.000000	1.000000	1.00000
$Subsidy_i$	16368.0	0.429008	0.494950	0.000000	0.000000	0.000000	1.000000	1.00000
$Year2014_t$	16368.0	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.00000
$Subsidy_i * Year2014_t$	16368.0	0.429008	0.494950	0.000000	0.000000	0.000000	1.000000	1.00000
$Income_i$	16368.0	5.839903	4.362678	1.381042	2.636535	4.519774	7.532957	25.48853
Δ_i	16368.0	1.839903	4.362678	-2.618958	-1.363465	0.519774	3.532957	21.48853
$Subsidy_i * \Delta_i$	16368.0	-0.622280	0.867030	-2.618958	-1.363465	0.000000	0.000000	0.00000

Model (4): $\mathbf{Y_i = B1 + B2{*}Subsidy_i + B3{*}\Delta_i + B4{*}(Subsidy{*}\Delta)_i + e_i}$

Y_i is whether or not the individual is insured (same as (3)). $Subsidy_i$ is a dummy for an individual's eligibility for the ACA subsidy (same as (1), (2), and (3)). Δ_i is the difference between the individual's income and the FPL cutoff for subsidy eligibility, which is 4.0. This can be written as $\Delta_i = (Income_i - 4.0)$. $(Subsidy*\Delta)_i$ represents the interaction term that is either 0 for a person whose income makes them ineligible for the subsidy or Δ_i for a person who is eligible for the subsidy based on income. Assuming the assumptions of a sharp regression discontinuity leveraging local linear regression hold, the causal treatment effect is estimated by B2. One important note about this model is that it leverages local linear regression, meaning that it is nonparametric and only uses data including people with an income within a .3 on either side of the FPL cutoff of 4 (ie. incomes between 3.7 and 4.3 FPL). This bandwidth is somewhat wide and constitutes an income change of about $7,500 on either side of the cutoff. This wide bandwidth is chosen due to needing sufficient data to create a model with reduced variance, although it should be noted that this can potentially introduce additional bias into the model. This was also run on data from 2015.

Figure 6: Summary Statistics for Nonparametric Regression Discontinuity Model (4) on Overall Adult Population Data

	count	mean	std	min	25%	50%	75%	max
Y_i	1508.0	0.934350	0.247751	0.000000	1.00000	1.000000	1.000000	1.000000
$Subsidy_i$	1508.0	0.468833	0.499193	0.000000	0.00000	0.000000	1.000000	1.000000
$Year2014_t$	1508.0	1.000000	0.000000	1.000000	1.00000	1.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	1508.0	0.468833	0.499193	0.000000	0.00000	0.000000	1.000000	1.000000
$Income_i$	1508.0	4.016956	0.204803	3.703704	3.82328	4.078165	4.248088	4.298434
Δ_i	1508.0	0.016956	0.204803	-0.296296	-0.17672	0.078165	0.248088	0.298434
$Subsidy_i * \Delta_i$	1508.0	-0.087228	0.106209	-0.296296	-0.17672	0.000000	0.000000	0.000000

Since both (3) and (4) are sharp regression discontinuity models, they have key assumptions that must be met in order for their B2s to estimate the true treatment effect. The most notable are as follows:

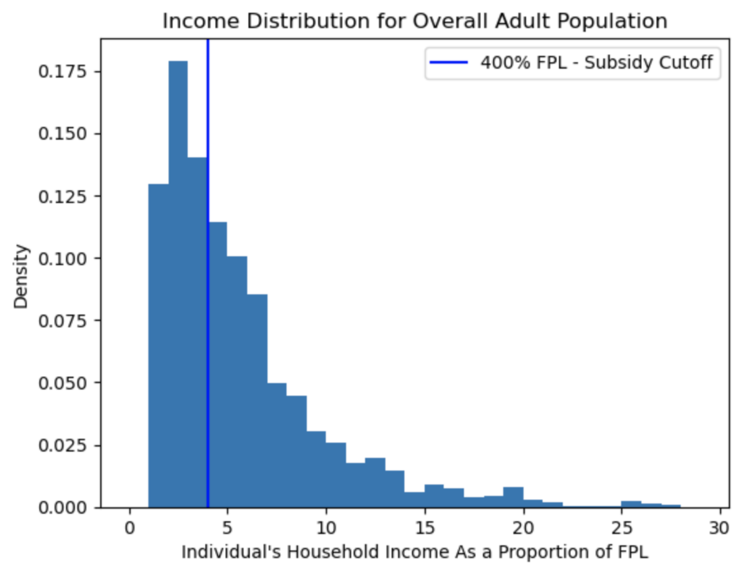
Assumption 1: *The running variable (Income) runs smoothly and continuously around the cut off (400% FPL).*

This is fulfilled as the running variable, an individual's income as a proportion of the FPL, runs continuously and smoothly around the cutoff of 400%.

Assumption 2: *Individuals close to the cutoff on both sides should be highly similar, on average, in all potentially relevant characteristics - observable and not observable.*

This is potentially unmet: given that people have some control over the inputs that drive their wage (jobs chosen, hours worked, etc), they may be manipulating their income to be on one side of the cut off, which creates selection bias. More concretely, a person may choose to work slightly less to stay under the 400% cutoff and earn the ACA subsidy if they would earn a similar amount by continuing to work. This would lead people with this idea to concentrate in the data with an income slightly below 400% and therefore shift the characteristics of those slightly above and below the cut off so they are not necessarily the same, thereby violating Assumption 2. However, based on the lack of bunching in the data distribution shown in the figure below, we do not see strong evidence for or against this idea.

Figure 7: Income Distribution for Overall Adult Population



Assumption 3 (only for the linear/parametric model): *The assumed relationship (in this case, linear) is the true relationship between the running and dependent variables.*

This is likely not fulfilled because the outcome variable is a boolean and the input is a continuous variable and a dummy variable. Since the data is not a clean split into the two separate categories of the boolean outcome on either side of the cut off, and the data has sufficient sample size to decently reflect the true underlying relationship, the underlying relationship is likely not a clean linear relationship. The non-parametric model avoids this assumption by only looking at data close to the cut off (local linear regression), so while the model itself is linear, it is not limited by assuming a global linear relationship and thus reduces the model's bias that would stem from including points far from the cutoff which deviate from the potentially erroneously-assumed linear relationship.

IV - Summary of Results for Overall Adult Population

The first model employed is the one described in the section above and marked as (1). It is the difference in differences model that includes all of the years in the dataset. The second model (2) is the same but restricted to the years 2014 (treatment year) and 2013 (control year). Their results are as follows:

Figure 8: Difference In Differences - Models (1) and (2) - on Overall Adult Population Results

	Dependent variable:Insured	
	(1)	(2)
Treatment Group	-0.094***	-0.081***
	(0.002)	(0.004)
Treatment Group and Year	0.042***	0.026***
	(0.004)	(0.006)
Treatment Year	0.007***	0.017***
	(0.002)	(0.004)
const	0.959***	0.953***
	(0.002)	(0.003)
Observations	80,460	32,132
R^2	0.026	0.022
Adjusted R^2	0.026	0.022
Residual Std. Error	0.256 (df=80456)	0.251 (df=32128)
F Statistic	720.381*** (df=3; 80456)	239.491*** (df=3; 32128)
Note:		*p<0.1; **p<0.05; ***p<0.01

Per the definitions in the empirical models and data section of this paper, we know that the estimated causal effect of healthcare subsidies on an individual's health insurance coverage status is the coefficient on "Treatment Group and Year." For model (1), this is .042 and is statistically significant at a 99% confidence level. Economically, this means that we can, per these results, claim that ACA healthcare subsidies causally increase an eligible individual's expected insurance coverage status. This said, the magnitude of this increase is incredibly small as it is represented by a .042 increase estimated for a binary variable. For model (2), this is .026 and is statistically significant at a 99% confidence level. This has the same economic interpretation as (1), but with an even smaller magnitude.

While these results are statistically and economically significant (although small), they must be taken with caution because the fulfillment of the assumptions necessary to make strong causal claims for difference in differences methods is questionable in this case, as shown in the previous section.

The third and fourth models in this paper are regression discontinuity models. The third is a parametric (linear) regression discontinuity and is described in the previous section as (3). The fourth is a (local linear) nonparametric regression discontinuity model that is described in the previous section as (4). The results are as follows:

Figure 9: Regression Discontinuity - Models (3) and (4) - on Overall Adult Population Results

	Dependent variable:Insured	
	(1)	(2)
Income as Prop FPL	0.002***	
	(0.001)	
Income as Prop. FPL Minus Cutoff Prop. FPL		-0.230*
		(0.119)
Interaction Term Local Lin Reg		0.153
		(0.171)
Treatment Group	-0.038***	-0.050
	(0.005)	(0.035)
const	0.947***	0.975***
	(0.005)	(0.025)
Observations	16,368	1,508
R^2	0.012	0.003
Adjusted R^2	0.012	0.001
Residual Std. Error	0.232 (df=16365)	0.248 (df=1504)
F Statistic	97.318*** (df=2; 16365)	1.546 (df=3; 1504)
Note:		*p<0.1; **p<0.05; ***p<0.01

As is described in the empirical models and data section of this paper, the estimated causal effect of healthcare subsidies on an individual's health insurance coverage status is the coefficient on "Treatment Group" for both models. For model (3), which is shown in column (1) of the previous figure, this is -.038 and is statistically significant at a 99% confidence level. Economically, these results claim that ACA healthcare subsidies causally decrease an eligible individual's expected insurance coverage status. This said, the magnitude of this decrease is very small (similar to the results for the difference in differences model). For model (4), which is shown in column (2) of the above figure, the causal effect is -.05 and is not statistically significant at any reasonable confidence level. This has the same economic interpretation as the first regression discontinuity model, however, it is not based on a statistically significant result so we cannot say that there is an economically significant result. Similarly, the magnitude of the result is very small, meaning that even if there is a projected effect, it is small regardless of the statistical significance.

While all of these results are statistically and economically significant except for one, it must be remembered that the assumptions necessary to make valid causal claims from regression discontinuity methods could be somewhat invalid for our dataset, as is described in the previous section of this paper.

V - Robustness Checks for the Overall Adult Population

Given the questions regarding the validity of the underlying assumptions of the models, as well as to be generally rigorous in analysis, this paper includes robustness checks. Specifically, it includes placebo tests on all of the models above.

To complete these placebo tests, the difference in differences models are adjusted to make the first treatment year earlier - in this case, it is 2012. By running the models with the new initial treatment year in 2012, the results can be compared to the results from the test with the actual treatment year. The results of the placebo tests provide an example of how the models would estimate the treatment if the treatment did not actually occur - indicating natural error/noise in the model. The difference in difference models (1) and (2) defined in the first data and model section of this paper are thus modified to

Placebo Model (1): $\mathbf{Y_{it} = B1 + B2*Subsidy_i + B3*Year2012_t + B4*(Subsidy*Year2012)_{it} + e_{it}}$

Placebo Model (2): Same equation as (1) but run on a dataset restricted to the placebo treatment year (2011) and the year prior (2012).

with the same variable definitions as the original models (besides the change of 2014_i to 2012_i). The results of the placebo tests for models (1) and (2) are in columns (1) and (2) of the following figure:

Figure 10: Difference In Differences - Placebo Models (1) and (2) - Overall Adult Population

	Dependent variable:Insured	
	(1)	(2)
Placebo Treatment Year	0.004	0.008*
	(0.003)	(0.004)
Placebo Treatment Year * Treatment Group	0.025***	-0.007
	(0.004)	(0.006)
Treatment Group	-0.098***	-0.098***
	(0.004)	(0.004)
const	0.959***	0.959***
	(0.003)	(0.003)
Observations	80,460	31,960
R^2	0.023	0.033
Adjusted R^2	0.023	0.033
Residual Std. Error	0.256 (df=80456)	0.272 (df=31956)
F Statistic	636.305*** (df=3; 80456)	362.446*** (df=3; 31956)
Note:		*p<0.1; **p<0.05; ***p<0.01

As is seen in these results, the placebo model (1) has an estimated causal effect of .025 and is significant at the 99% confidence level. Economically, this means that the first placebo model estimates the causal effect of subsidies on an individual's health insurance coverage status is an increase of 025. Given that this result is for the placebo test and is similar to the actual results from the first model (.042 and statistically significant), we must view the results of the original model with extra skepticism. For placebo model (2), the estimated causal effect is -.007. This is not statistically significant. Since the placebo results are not statistically significant, this adds confidence in the estimate of causal effects from the actual model (2) (which is .026).

Placebo models for both regression discontinuity models were also run for the same purposes. They take the same form as the original models (3) and (4) but are run on data from 2012 which is before the ACA subsidies were enacted in California.

Placebo Model (3): $\mathbf{Y_i = B1 + B2{*}Subsidy_i + B3{*}Income_i + e_i}$

Placebo Model (4): $\mathbf{Y_i = B1 + B2{*}Subsidy_i + B3{*}\Delta_i + B4{*}(Subsidy{*}\Delta)_i + e_i}$

The results from placebo model (3) are shown in column (1) and placebo model (4) in column (2) of the figure below.

Figure 11: Regression Discontinuity - Placebo Models (3) and (4) - Overall Adult Population

	Dependent variable:Insured	
	(1)	(2)
Income as Prop FPL	0.004***	
	(0.001)	
Income as Prop. FPL Minus Cutoff Prop. FPL		0.034
		(0.111)
Interaction Term Local Lin Reg		0.013
		(0.153)
Treatment Group	-0.080***	0.005
	(0.006)	(0.027)
const	0.932***	0.933***
	(0.007)	(0.022)
Observations	14,605	1,060
R^2	0.038	0.000
Adjusted R^2	0.038	-0.002
Residual Std. Error	0.270 (df=14602)	0.247 (df=1056)
F Statistic	287.140*** (df=2; 14602)	0.155 (df=3; 1056)
Note:	*p<0.1; **p<0.05; ***p<0.01	

As is seen in the results, the estimated causal effect from the placebo model (3) on an individual from the overall adult population is -.08 with statistical significance at a 99% confidence level. This compares to the -.038 with the same statistical significance estimate from the actual model. Given the significance and large magnitude of the placebo results (relative to the actual results), the actual model and its results must be assessed with added skepticism. For placebo model (4), the causal estimate for an individual in the overall adult population is .005. This is not statistically significant, so we do not need to alter our interpretation of the results from the actual model (4).

V - Empirical Models and Data for the African American Adult Population

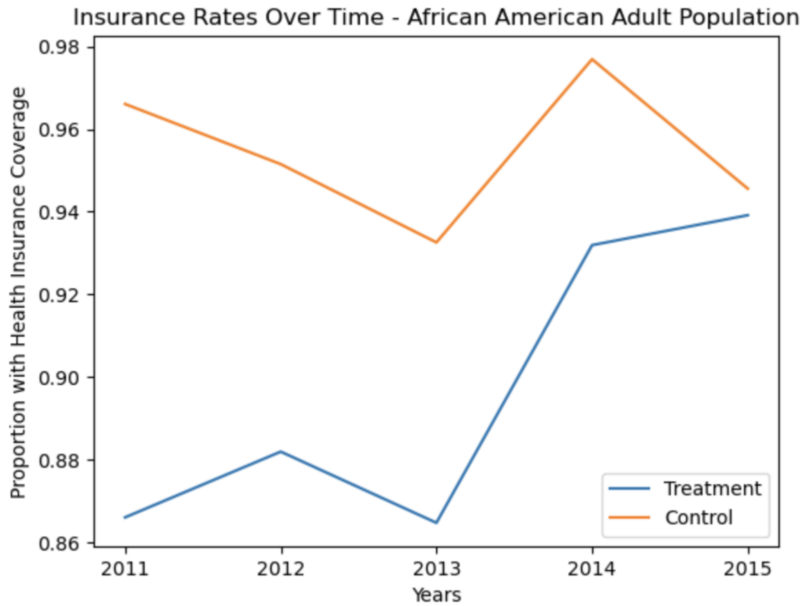
For this section, the same dataset as before is employed, however, it is filtered to only include African Americans (so the data is Californian adults that are African American) to generate an exclusively-African American dataset. This filtering reduces the sample sizes significantly, but the results sizes are still adequate for the empirical models. The sample sizes are as follows:

Figure 12: Sample Sizes Per Year for the African American Adult Population

Years	AA Counts	AA Treatment Counts	AA Control Counts
2011	771	388	383
2012	649	381	268
2013	740	399	341
2014	626	323	303
2015	901	460	441

For the analysis that is run on data that exclusively contains African American adults, the methods are the same as those contained in the previous methods section (III). This means that the empirical technicals, equations, key assumptions, and concerns regarding the fulfillment of those key assumptions are the same. The racial filter creates two notable potential shifts in the fulfillment of the key assumptions: The first potential shift in the validity of key assumptions is in regard to the parallel trends assumption for the difference in differences models. As described in the overall models and data section, this assumption is paramount for validly interpreting the models' results as causal and significant. Similar to the results for the overall group, the pre-trends for the African American data seem to questionably fulfill this assumption and thus should be considered when evaluating the economic significance and validity of the results. The results pre-trends for the African American population are as follows:

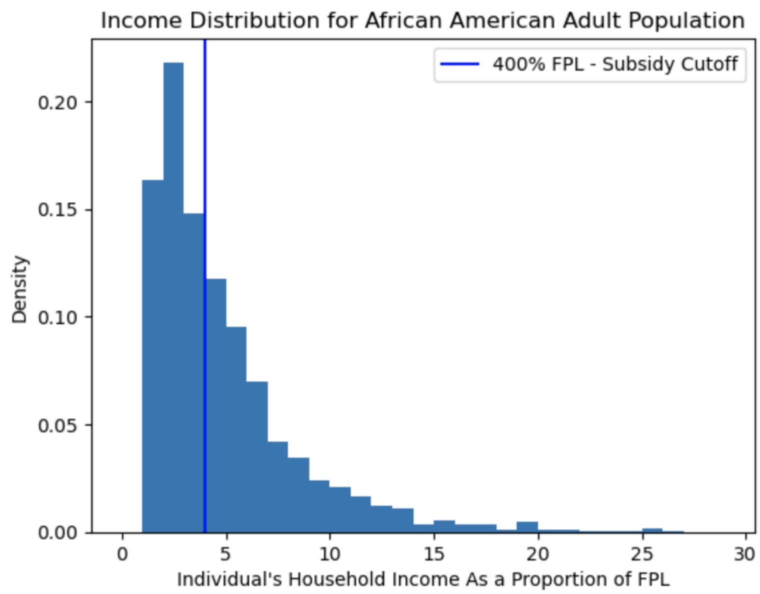
Figure 13 Percentage of Insurance Coverage Per Year - African American Population



Note: for interpretability, the graph shows aggregate insurance rates, not an individual's insurance status. While this is different from individual insurance coverage status, the boolean nature of the status lends well to aggregation, making the visualization a quality representation of pretrends.

The second potential shift in the validity of key assumptions regards the manipulation of income described previously, where individuals may alter their income to be below the cutoff and have subsidy eligibility. Similarly to the previous models and data section, strong evidence for or against this phenomenon is not shown in the data, as can be seen in the following figure.

Figure 14: Income Distribution for the Adult African Americans Population



The summary statistics, however, are not the same since non-African Americans have been excluded. The summary statistics for the models are as follows:

Figure 15: Summary Statistics for DID Model (1) on African American-Only Data

	count	mean	std	min	25%	50%	75%	max
Y_{it}	3687.0	0.923786	0.265376	0.000000	1.000000	1.000000	1.000000	1.000000
$Subsidy_i$	3687.0	0.529156	0.499217	0.000000	0.000000	1.000000	1.000000	1.000000
$Year2014_t$	3687.0	0.414158	0.492643	0.000000	0.000000	0.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	3687.0	0.212368	0.409039	0.000000	0.000000	0.000000	0.000000	1.000000
$Income_i$	3687.0	4.856229	3.584863	1.381042	2.313625	3.766478	6.117767	26.109661
Δ_i	3687.0	0.856229	3.584863	-2.618958	-1.686375	-0.233522	2.117767	22.109661
$Subsidy_i * \Delta_i$	3687.0	-0.794657	0.913239	-2.618958	-1.686375	-0.233522	0.000000	0.000000

Figure 16: Summary Statistics for Reduced Time Frame DID Model (2) on African American-only Data

	count	mean	std	min	25%	50%	75%	max
Y_{it}	1366.0	0.922401	0.267637	0.000000	1.000000	1.000000	1.000000	1.000000
$Subsidy_i$	1366.0	0.528551	0.499367	0.000000	0.000000	1.000000	1.000000	1.000000
$Year2014_t$	1366.0	0.458272	0.498438	0.000000	0.000000	0.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	1366.0	0.236457	0.425061	0.000000	0.000000	0.000000	0.000000	1.000000
$Income_i$	1366.0	4.897256	3.560955	1.392515	2.436902	3.814367	6.169666	26.109661
Δ_i	1366.0	0.897256	3.560955	-2.607485	-1.563098	-0.185633	2.169666	22.109661
$Subsidy_i * \Delta_i$	1366.0	-0.764676	0.891968	-2.607485	-1.563098	-0.185633	0.000000	0.000000

Figure 17: Summary Statistics for Parametric Regression Discontinuity Model (3) on African American-Only Data

	count	mean	std	min	25%	50%	75%	max
Y_i	901.0	0.942286	0.233331	0.000000	1.000000	1.000000	1.000000	1.00000
$Subsidy_i$	901.0	0.510544	0.500166	0.000000	0.000000	1.000000	1.000000	1.00000
$Year2014_t$	901.0	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.00000
$Subsidy_i * Year2014_t$	901.0	0.510544	0.500166	0.000000	0.000000	1.000000	1.000000	1.00000
$Income_i$	901.0	5.013720	3.787666	1.381042	2.322662	3.857349	6.372133	25.48853
Δ_i	901.0	1.013720	3.787666	-2.618958	-1.677338	-0.142651	2.372133	21.48853
$Subsidy_i * \Delta_i$	901.0	-0.781661	0.922052	-2.618958	-1.677338	-0.142651	0.000000	0.00000

Figure 18: Summary Statistics for Nonparametric Regression Discontinuity Model (4) on African American-Only Data

	count	mean	std	min	25%	50%	75%	max
Y_i	88.0	0.954545	0.209493	0.000000	1.00000	1.000000	1.000000	1.000000
$Subsidy_i$	88.0	0.386364	0.489706	0.000000	0.00000	0.000000	1.000000	1.000000
$Year2014_t$	88.0	1.000000	0.000000	1.000000	1.00000	1.000000	1.000000	1.000000
$Subsidy_i * Year2014_t$	88.0	0.386364	0.489706	0.000000	0.00000	0.000000	1.000000	1.000000
$Income_i$	88.0	4.063410	0.190222	3.766478	3.82328	4.123711	4.248088	4.268675
Δ_i	88.0	0.063410	0.190222	-0.233522	-0.17672	0.123711	0.248088	0.268675
$Subsidy_i * \Delta_i$	88.0	-0.061095	0.089776	-0.233522	-0.17672	0.000000	0.000000	0.000000

As can be seen from the summary stats, the sample sizes range from 3687 to 1366 to 901 to 88 for the experiments. While all of the sizes should be sufficient for analysis, the results from model (4) should be considered with caution due to the relatively low sample size.

VI - Summary of Results for African American Adult Population

As mentioned previously, the difference in difference models (1) and (2) were also run on a dataset restricted to African American adults in order to generate results that can be independently evaluated and compared to the results from the overall population. Their results are shown in the following figure with model (1) in column (1) and model (2) in column (2):

Figure 19: Difference in Difference - Models (1) and (2) - Results for the Adult African American Population

	Dependent variable:Insured	
	(1)	(2)
Treatment Group	-0.080***	-0.068***
	(0.011)	(0.020)
Treatment Group and Year	0.058***	0.023
	(0.018)	(0.029)
Treatment Year	0.008	0.044**
	(0.013)	(0.021)
const	0.951***	0.933***
	(0.008)	(0.014)
Observations	3,687	1,366
R^2	0.019	0.023
Adjusted R^2	0.019	0.021
Residual Std. Error	0.263 (df=3683)	0.265 (df=1362)
F Statistic	24.162*** (df=3; 3683)	10.909*** (df=3; 1362)
Note:		*p<0.1; **p<0.05; ***p<0.01

As described previously, the estimated causal effect is the coefficient on the Treatment Group and Year variable for both difference in differences models. For model (1), this is .058 and is statistically significant at the 99% confidence level. Economically, this means that the model implies that eligibility for available ACA healthcare subsidies causally increases the expectation of an African American individual's health insurance coverage status by .058. This is a highly minute increase as the healthcare coverage status variable is a boolean (an increase of 1 would mean that it fully causes coverage). For model (2), the estimated causal effect is .023, but it is not statistically significant at any reasonable confidence level. This would have the same economic interpretation as that of model (1), however, the magnitude is very small and it is not statistically significant. This means model (2) does not have results that are statistically or economically significant to determine a causal effect of ACA subsidies on African American individuals' health insurance coverage statuses.

As is discussed in the "III - Empirical Models and Data for the Overall Adult Population" section of this paper, there are valid concerns about the assumptions behind this model being fulfilled. Additionally, the reduced dataset also introduces concerns about a smaller sample size. This means the results must be interpreted with increased skepticism and attention.

The regression discontinuity models were also run on exclusively African American data to empower standalone analysis and comparison. In the following figure, the results from parametric regression discontinuity (model (3)) are shown in column (1) and the results from the nonparametric regression discontinuity model (model (4)) are shown in column (2).

Figure 20: Regression Discontinuity - Models (3) and (4) - Results for the Adult African American Population

	Dependent variable:Insured	
	(1)	(2)
Income as Prop FPL	0.001	
	(0.003)	
Income as Prop. FPL Minus Cutoff Prop. FPL		-0.203
		(0.426)
Interaction Term Local Lin Reg		0.307
		(0.658)
Treatment Group	-0.000	0.001
	(0.021)	(0.126)
const	0.936***	0.986***
	(0.024)	(0.091)
Observations	901	88
R^2	0.000	0.007
Adjusted R^2	-0.002	-0.029
Residual Std. Error	0.234 (df=898)	0.212 (df=84)
F Statistic	0.181 (df=2; 898)	0.195 (df=3; 84)
Note:		*p<0.1; **p<0.05; ***p<0.01

As previously detailed the estimated causal effect is the coefficient on the Treatment Group variable for both regression discontinuity models (models (3) and (4)). For the parametric model (model (3)), this takes value -.000 and is not statistically, and thus not economically, significant at any legitimate confidence level. The nonparametric model (model (4)) has a similar result of .001 and no statistical significance. The resulting takeaways and logic are the same as for model (3) - nothing with meaningful statistical or economic significance. Therefore, models (3) and (4) do not provide useful evidence for the causal effects of healthcare subsidies on African American individuals' health insurance coverage statuses.

As is mentioned in previous sections, these results must be interpreted while remembering that there are potential violations of the necessary assumptions to the regression discontinuity models. In addition, the sample sizes for these models, especially model (4), are small enough to potentially have adverse effects on the validity of our results.

VII - Robustness Checks for the African American Adult Population

The same placebo models run on the overall adult population were also run on the data exclusive to African Americans. As a refresher, these models are the regular difference in differences models with the initial treatment year shifted to 2012, as well as the regression discontinuity models run on data from the year 2012.

The results of the placebo difference in differences models (1) and (2) are shown in columns (1) and (2) of the results below:

Figure 21: Difference in Difference - Placebo Models (1) and (2) - Results for the Adult African American Population

	Dependent variable:Insured	
	(1)	(2)
Placebo Treatment Year	-0.016	-0.015
	(0.015)	(0.022)
Placebo Treatment Year * Treatment Group	0.054**	0.030
	(0.021)	(0.030)
Treatment Group	-0.100***	-0.100***
	(0.019)	(0.020)
const	0.966***	0.966***
	(0.013)	(0.014)
Observations	3,687	1,420
R^2	0.014	0.024
Adjusted R^2	0.013	0.022
Residual Std. Error	0.264 (df=3683)	0.278 (df=1416)
F Statistic	16.858*** (df=3; 3683)	11.630*** (df=3; 1416)
Note:		*p<0.1; **p<0.05; ***p<0.01

The results of placebo model (1) applied to the exclusively African American data indicate a causal effect of .054 which is significant at a 95% confidence level. This means that the estimated causal effect of the actual model (.058, statistically significant at the 99% confidence level) is likely partially due to the nature of the model and the data, not the actual causal effect. As such, these results from placebo model (1) make the results of applying the actual model (1) on the exclusively African American data more questionable. The results of placebo model (2) applied to the exclusively African American data indicate a causal effect of .03 which is not statistically at any reasonable confidence level. Although this result and the result from the actual model 2 are similar (.03 vs .023), the interpretation is not really altered since the results were already not statistically, and thus economically, significant.

The results for the placebo regression discontinuity models (3) and (4) are as in columns (1) and (2) of the following:

Figure 22: Regression Discontinuity - Placebo Models (3) and (4) - Results for the Adult African American Population

	Dependent variable:Insured	
	(1)	(2)
Income as Prop FPL	0.010**	
	(0.005)	
Income as Prop. FPL Minus Cutoff Prop. FPL		-0.898*
		(0.491)
Interaction Term Local Lin Reg		0.898
		(0.733)
Treatment Group	-0.019	-0.003
	(0.032)	(0.138)
const	0.875***	1.003***
	(0.038)	(0.096)
Observations	649	43
R^2	0.022	0.151
Adjusted R^2	0.019	0.086
Residual Std. Error	0.283 (df=646)	0.246 (df=39)
F Statistic	7.310*** (df=2; 646)	2.321* (df=3; 39)
Note:		*p<0.1; **p<0.05; ***p<0.01

As is seen in these results, the estimated causal effect is -.019 for placebo model (3) and is not statistically significant. This does not alter the interpretation of the result from the actual model (0 causal effect) since both are statistically insignificant. The causal estimate from placebo model (4) being applied to the African American dataset is -.003. Similarly to the interpretation of the results for placebo model (3) with the African American dataset, the results from the placebo model (4) when applied to the African American dataset do not meaningfully alter the interpretation of the actual model 4's results since both were already statistically and economically insignificant.

VIII - Comparison of Results for Overall Adult Population and African American Adult Population

Overall, the results of the models run on the overall dataset and the African American dataset are similar. The causal estimates of interest from the difference in differences models are, for example, all within .02 of each other. Given that the outcome variable is a boolean and is out of 1, this indicates that the results are very similar. While the second difference in differences model does not have statistically significant results for its causal estimate when run on the African American data, unlike the estimates for the overall population and model (1)'s estimate for the African American population, this is likely due to the reduced sample size of the African American-only data. The results are still highly similar.

The comparison of the results on the regression discontinuity models being applied to the overall and African American datasets have similar implications. Again, the results are all within .05 of each other. However, it is worth noting that only the overall data had a statistically significant estimate for the causal effect for these two models with an estimate of -.038. While this is significant at a 99% confidence level, we see a similar result from the associated placebo test. Also, the reduced sample size and small magnitude of both the African American and overall results make the difference between these results practically negligible.

Due to the similarity of the results across the overall and African American dataset, along with the overall minute nature of all of the causal estimates, the results do not imply a significant difference in the causal effects of subsidies on individuals' insurance coverage statuses between the overall and African American populations.

IX - Discussion and Conclusion

Overall, the results shown in this paper are mixed. While the goal of this paper was to identify and quantify the causal effect of health insurance subsidies on individuals' health insurance coverage statuses, both among the general and overall populations, the results ultimately did not prove a notable causal effect with any statistical significance. In fact, the magnitudes (absolute values) of the estimates for the causal effect of health insurance subsidies on an individual's insurance coverage status are all less .058. Since the endogenous variable, insurance coverage status, is a boolean (takes value 1 or 0), this effect is incredibly minute and limits the economic significance of all of the results.

Additionally, some of the results determined a minute negative effect (all but one of the results from the regression discontinuity models), yet others showed a minute positive effect (all of the results from the difference in differences models). Given that the actual causal effect cannot simultaneously be positive and negative yet there are statistically significant results with both positive and negative signs, it is implied that there is an issue regarding the dataset (which is unlikely, but possible due to the size, collection method, etc.), or there are violations of the assumptions necessary for the models that ended up being significant in the results. Together, these results with opposite polarities provide conflicting results and interpretations of the causal effects of health insurance subsidies on individuals' health insurance coverage statuses; however, as mentioned above, all of the results are so small in magnitude that they can be thought of as practically economically insignificant.

Another challenge to the results of the models is the results of the placebo tests - all but one of the actual models' results with statistical significance had similar results in their associated placebo tests. This means that those statistically significant results may have been more due to issues and/or randomness within the models and data employed than the actual underlying relationship, further casting the (already small enough to be economically negligible) estimated causal effects into doubt.

The most concrete result from the analysis in this paper is a statistically significant result that indicated that health insurance subsidies increase individuals' expected health insurance coverage statuses by .026 for people in the overall adult population. This result (from model (2)) is the only result that is statistically significant and has an associated placebo test that did not have significant results (which may be been due to sample sizes, randomness, etc., so additional analysis is still be needed to be certain in interpreting a precise causal effect). Regardless, this result is so small the economic significance is negligible; therefore, this paper, and its unique combination of data and empirical methods, were unable to prove any economically-significant, large causal effects of subsidies on health insurance coverage statuses. Additionally, the results for running analysis on the African American and overall population dataset were very similar, and, while this would indicate that there is not much of a disparity in the investigated causal effect depending on race, the results of the original analysis are small and uncertain enough to cast skepticism on this conclusion and dictate that this question should be addressed in papers employing different methods and/or data.

While this paper was not successful in identifying a concrete, strong causal effect of health insurance subsidies on individuals' health insurance coverage statuses, it does inspire many ideas for follow-up economic analysis. First, there is a very natural follow-up: re-running the analysis with data that certainly and fully fulfills the assumptions necessary for the employed models. (As mentioned in the models and data section, it is uncertain if the employed data fulfills the parallel trends assumption, SUTVA, the similarity on opposite sides of the cutoff assumption, and the underlying linear relationship assumption.) While the current data was likely close to fulfilling or did fulfill most of these assumptions, having a dataset that certainly fulfills all these assumptions would empower analysis with results that are valid with increased certainty.

An additional provoking area for future analysis is as follows: Given the problems found with validating the assumptions necessary for a successful regression discontinuity, other models and data could be explored to attempt to find a more certain effect. For example, data that includes the amount of insurance purchased by individuals could depict a more linear relationship between income, subsidy eligibility, and the amount of insurance purchased. This would make the parametric regression discontinuity model more useful and potentially generate some interesting results. Similarly, a logit model could be employed on the current data and is potentially a better fit for the underlying relationship (a running variable, a dummy variable, and a boolean outcome variable).

Finally, another exciting area for future analysis lies in altering the data sources and filtering employed in this paper. For example, restricting data to those who did not have group health insurance would allow for analysis of a group that is more frequently a target of healthcare subsidies and seems logically more likely to respond to them. This would, in theory, present a great opportunity to identify a significant, causal increase in expected insurance coverage statuses created by healthcare subsidies. Identifying a dataset that has higher response rates could also reduce the highly valid concerns about volunteer bias that are present when using the CHIS dataset. In addition, the mentioned analyses could be run on data that includes and delineates individuals' statuses in other demographic subgroups, such as children, adolescents, racial groups, people with different types of careers, etc. This would empower rigorous analysis to determine and compare, across many different demographics, the causal effects of healthcare subsidies on individuals' health insurance coverage statuses.

References

Buchmueller, Thomas C., Zachary M. Levinson, Helen G. Levy, and Barbara L. Wolfe. 2016. "Effect of the Affordable Care Act on Racial and Ethnic Disparities in Health Insurance Coverage." *American Journal of Public Health*, vol. 106, no. 8. doi:10.2105/AJPH.2016.303155.

Card, David, and Alan Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review*, vol. 90, no. 5. https://davidcard.berkeley.edu/papers/njmin-aer.pdf.

Courtemanche, Charles, James Marton, Benjamin Ukert, Aaron Yelowitz, Daniela Zapata. "Early Impacts of the Affordable Care Act on Health Insurance Coverage in Medicaid Expansion and Non-Expansion States." *Journal of Policy Analysis and Management*, vol. 36, no. 1. doi.org/10.1002/pam.21961.

Duggan, Mark, Gopi Shah Goda, and Gina Li. 2021. "The Effects of the Affordable Care Act on the Near Elderly: Evidence for Health Insurance Coverage and Labor Market Outcomes." *Tax Policy and the Economy*, vol. 35. https://doi.org/10.1086/713496.

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers. 2017. "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act." *Journal of Health Economics,* vol. 53, May 2017, pp. 72-86. https://doi.org/10.1016/j.jhealeco.2017.02.004.

Goldman, Anna L., Danny McCormick, Jennifer S. Haas, and Benjamin D. Sommers. 2018. "Effects Of The ACA's Health Insurance Marketplaces On The Previously Uninsured: A Quasi-Experimental Analysis." *Health Affairs,* vol. 25, no. 4. doi.org/10.1377/hlthaff.2017.1390.

Kettlewell, Nathan, Olena Stavrunova, and Oleg Yerokhin. 2018. “Premium subsidies and demand for private health insurance: results from a regression discontinuity design.” *Applied Economics Letters*, vol. 25, no. 2. https://doi.org/10.1080/13504851.2017.1299094.

Marquis, Susan, Melinda Beeuwkes Buntin, José J. Escarce, Kanika Kapur, and Jill M. Yegian. 2004. “Subsidies and the Demand for Individual Health Insurance in California.” *Health Services Research*, vol. 39, no. 5. https://doi.org/10.1111/j.1475-6773.2004.00303.x.

McWilliams, J. Michael. 2009. “Health consequences of uninsurance among adults in the United States: recent evidence and implications.” *The Milbank Quarterly*, vol. 87, no. 2. DOI: 10.1111/j.1468-0009.2009.00564.x.

Wallace, Jacob, Karen Jiang, and Paul Goldsmith-Pinkham. 2021. “Changes in Racial and Ethnic Disparities in Access to Care and Health Among US Adults at Age 65 Years.” *Jama Internal Medicine*, vol. 181, no. 9. doi:10.1001/jamainternmed.2021.3922.

Appendix

Appendix Figure 1: Overall Population DID - Models (1), (2), Placebo (1), and Placebo (2) - shown in order

				Dependent variable:Insured
	(1)	(2)	(3)	(4)
Placebo Treatment Year			0.004	0.008*
			(0.003)	(0.004)
Placebo Treatment Year * Treatment Group			0.025***	-0.007
			(0.004)	(0.006)
Treatment Group	-0.094***	-0.081***	-0.098***	-0.098***
	(0.002)	(0.004)	(0.004)	(0.004)
Treatment Group and Year	0.042***	0.026***		
	(0.004)	(0.006)		
Treatment Year	0.007***	0.017***		
	(0.002)	(0.004)		
const	0.959***	0.953***	0.959***	0.959***
	(0.002)	(0.003)	(0.003)	(0.003)
Observations	80,460	32,132	80,460	31,960
R^2	0.026	0.022	0.023	0.033
Adjusted R^2	0.026	0.022	0.023	0.033
Residual Std. Error	0.256 (df=80456)	0.251 (df=32128)	0.256 (df=80456)	0.272 (df=31956)
F Statistic	720.381*** (df=3; 80456)	239.491*** (df=3; 32128)	636.305*** (df=3; 80456)	362.446*** (df=3; 31956)
Note:				*p<0.1; **p<0.05; ***p<0.01

Appendix Figure 2: Overall Population Regression Discontinuity - Models (3), (4), Placebo (3), and Placebo (4) - shown in order

				Dependent variable:Insured
	(1)	(2)	(3)	(4)
Income as Prop FPL	0.002***		0.004***	
	(0.001)		(0.001)	
Income as Prop. FPL Minus Cutoff Prop. FPL		-0.230*		0.034
		(0.119)		(0.111)
Interaction Term Local Lin Reg		0.153		0.013
		(0.171)		(0.153)
Treatment Group	-0.038***	-0.050	-0.080***	0.005
	(0.005)	(0.035)	(0.006)	(0.027)
const	0.947***	0.975***	0.932***	0.933***
	(0.005)	(0.025)	(0.007)	(0.022)
Observations	16,368	1,508	14,605	1,060
R^2	0.012	0.003	0.038	0.000
Adjusted R^2	0.012	0.001	0.038	-0.002
Residual Std. Error	0.232 (df=16365)	0.248 (df=1504)	0.270 (df=14602)	0.247 (df=1056)
F Statistic	97.318*** (df=2; 16365)	1.546 (df=3; 1504)	287.140*** (df=2; 14602)	0.155 (df=3; 1056)
Note:				*p<0.1; **p<0.05; ***p<0.01

Appendix Figure 3: AA DID - Models (1), (2), Placebo (1), and Placebo (2) - shown in order

				Dependent variable:Insured
	(1)	(2)	(3)	(4)
Placebo Treatment Year			-0.016	-0.015
			(0.015)	(0.022)
Placebo Treatment Year * Treatment Group			0.054**	0.030
			(0.021)	(0.030)
Treatment Group	-0.080***	-0.068***	-0.100***	-0.100***
	(0.011)	(0.020)	(0.019)	(0.020)
Treatment Group and Year	0.058***	0.023		
	(0.018)	(0.029)		
Treatment Year	0.008	0.044**		
	(0.013)	(0.021)		
const	0.951***	0.933***	0.966***	0.966***
	(0.008)	(0.014)	(0.013)	(0.014)
Observations	3,687	1,366	3,687	1,420
R^2	0.019	0.023	0.014	0.024
Adjusted R^2	0.019	0.021	0.013	0.022
Residual Std. Error	0.263 (df=3683)	0.265 (df=1362)	0.264 (df=3683)	0.278 (df=1416)
F Statistic	24.162*** (df=3; 3683)	10.909*** (df=3; 1362)	16.858*** (df=3; 3683)	11.630*** (df=3; 1416)
Note:				*p<0.1; **p<0.05; ***p<0.01

Appendix Figure 4: AA Regression Discontinuity - Models (3), (4), Placebo (3), and Placebo (4) - shown in order

				Dependent variable:Insured
	(1)	(2)	(3)	(4)
Income as Prop FPL	0.001		0.010**	
	(0.003)		(0.005)	
Income as Prop. FPL Minus Cutoff Prop. FPL		-0.203		-0.898*
		(0.426)		(0.491)
Interaction Term Local Lin Reg		0.307		0.898
		(0.658)		(0.733)
Treatment Group	-0.000	0.001	-0.019	-0.003
	(0.021)	(0.126)	(0.032)	(0.138)
const	0.936***	0.986***	0.875***	1.003***
	(0.024)	(0.091)	(0.038)	(0.096)
Observations	901	88	649	43
R^2	0.000	0.007	0.022	0.151
Adjusted R^2	-0.002	-0.029	0.019	0.086
Residual Std. Error	0.234 (df=898)	0.212 (df=84)	0.283 (df=646)	0.246 (df=39)
F Statistic	0.181 (df=2; 898)	0.195 (df=3; 84)	7.310*** (df=2; 646)	2.321* (df=3; 39)
Note:				*p<0.1; **p<0.05; ***p<0.01

Appendix Figure 5: Results Characteristics Table

Model	*Statistically Significant*	*Sign*	*Placebo Statistically Significant*	*Placebo Sign*
Model (1) - Overall Data	x	+	x	+
Model (1) - African American Data	x	+	x	+
Model (2) - Overall Data	x	+		-
Model (2) - African American Data		+		+
Model (3) - Overall Data	x	-	x	-
Model (3) - African American Data		-		-
Model (4) - Overall Data		-		+
Model (4) - African American Data		+		-

Note: an "x" indicates that the criterion is met for that model. A "+" indicates a positive sign and a "-" indicates a negative sign.